

# Generalized subspace subcodes with application in cryptology

Thierry P. BERGER, Cheikh Thiécoumba GUEYE and Jean Belo KLAMTI

Cheikh Thiécoumba GUEYE and Jean Belo KLAMTI are with Université Cheikh Anta Diop, Faculté des Sciences et Techniques, DMI, LACGAA, Dakar, Sénégal, cheikh.gueye@ucad.edu.sn and jeanbelo.klamti@ucad.edu.sn

Thierry P. Berger is with XLIM (UMR CNRS 6172), Université de Limoges, 123 avenue A. Thomas, 87060 Limoges Cedex, France, thierry.berger@unilim.fr

## Abstract

Most of the codes that have an algebraic decoding algorithm are derived from the Reed Solomon codes. They are obtained by taking equivalent codes, for example the generalized Reed Solomon codes, or by using the so-called subfield subcode method, which leads to Alternant codes and Goppa codes over the underlying prime field, or over some intermediate subfield. The main advantages of these constructions is to preserve both the minimum distance and the decoding algorithm of the underlying Reed Solomon code. In this paper, we propose a generalization of the subfield subcode construction by introducing the notion of subspace subcodes and a generalization of the equivalence of codes which leads to the notion of generalized subspace subcodes. When the dimension of the selected subspaces is equal to one, we show that our approach gives exactly the family of the codes obtained by equivalence and subfield subcode technique. However, our approach highlights the links between the subfield subcode of a code defined over an extension field and the operation of puncturing the  $q$ -ary image of this code. When the dimension of the subspaces is greater than one, we obtain codes whose alphabet is no longer a finite field, but a set of  $r$ -uples. We explain why these codes are practically as efficient for applications as the codes defined on an extension of degree  $r$ . In addition, they make it possible to obtain decodable codes over a large alphabet having parameters previously inaccessible. As an application, we give some examples that can be used in public key cryptosystems such as McEliece.

## Index Terms

Linear code, Shortened code, Punctured code, Subfield subcodes, Reed Solomon codes, Alternant codes,  $q$ -ary image.

## I. INTRODUCTION

The McEliece cryptosystem is the most known and oldest code-based cryptographic protocol. An important part of its security is based on the use of codes that seem random and possess an effective error correction algorithm. In its original paper, R. McEliece proposed the use of binary Goppa codes. This class is a subclass of Alternant codes, which are themselves subcodes on the binary field of Generalized Reed-Solomon codes. This construction makes it possible to easily decode errors, provides a good minimum distance and effectively mask the underlying algebraic structure.

The main problem with this protocol is the size of the secret key. There are several ways to reduce the size of keys. One of these is the use of codes with a large automorphism group, typically quasi-cyclic (QC), quasi-dyadic (QD), or quasi-monoidic (QM) matrices [2], [3], [9], [12], [13], [14], [17].

Another approach is to use subfield subcodes over a subfield of great size. The variant based on the subfield subcodes introduced by Berger *et al.* [3] was attacked by Wieschebrink [20]. Recently Faugère *et al.* proposed two attacks respectively a structural attack and an algebraic attack against the McEliece schemes with compact keys [6], [7].

In this paper, we introduce a new construction of subfield subcodes called *Generalized Subfield Subcodes* and we prove that the *Generalized Subfield Subcodes* of Reed-Solomon are exactly alternant codes. The approach developed for the *Generalized Subfield Subcodes* leads to a second construction called *Generalized Subspace Subcodes* which is a promising research direction for both coding theory and hiding the structure of a code.

This paper is organized as following: in Section II we give some definitions in coding theory. In Section III we introduce the shortened  $q$ -ary images of a code and give the link between subfield subcodes and shortened codes. In Section IV we present the first attempt at generalization of subfield subcodes namely the *Generalized Subfield Subcodes* and we show that the codes introduced in Section III-A can also be constructed using a known method to construct alternant codes. In Section V we introduce *Subspace Subcodes*, which is a new class of additive block codes. We generalize this first class to obtain another class of block additive codes named *Generalized Subspace Subcodes*. For this second class we proposed an algorithm which allows us to compute its generator matrix. In addition we give some examples and directions for their application in transmission and cryptology.

## II. PRELIMINARIES

### A. Linear code

Let  $\mathbb{F}_{q^m}$  be an arbitrary finite field. A linear code  $\mathcal{C}$  of length  $n$  and dimension  $k$  is a vector subspace of  $\mathbb{F}_{q^m}^n$  of dimension  $k$ . A vector  $x \in \mathbb{F}_{q^m}^n$  is called word and a vector  $x \in \mathcal{C}$  is called codeword.

The Hamming distance between two words  $x$  and  $y$  denoted by  $d(x, y)$  is the number of positions on which they differ. The Hamming distance of a code  $\mathcal{C}$  denoted by  $d$  is the minimal Hamming distance between any two different codewords.

The Hamming weight of a word  $x \in \mathbb{F}_{q^m}^n$  denoted by  $wt(x)$  is the number of its nonzero coordinates. In the case of a linear code the minimal Hamming distance of a code is equal to the minimal Hamming weight of its nonzero codewords.

A linear code  $\mathcal{C}$  over an arbitrary finite field  $\mathbb{F}_{q^m}$  is called  $\mathbb{F}_{q^m}$ -linear code. If its length is  $n$ , its dimension is  $k$  and its minimal Hamming distance is  $d$  we call this code a  $[n, k, d]$   $\mathbb{F}_{q^m}$ -linear code. A linear code  $\mathcal{C}$  over an arbitrary finite field  $\mathbb{F}_{q^m}$  is usually specified by a full-rank matrix  $\mathcal{G} \in \mathbb{F}_{q^m}^{k \times n}$  called *generator matrix* of  $\mathcal{C}$ , whose rows span the code. Thus,  $\mathcal{C} = \{x\mathcal{G} : x \in \mathbb{F}_{q^m}^k\}$ . A linear code can be also defined by the right kernel of a matrix  $\mathbf{H}$  called *parity-check matrix* of  $\mathcal{C}$  as follows:

$$\mathcal{C} = \{x \in \mathbb{F}_{q^m}^k \text{ s.t. } \mathbf{H}x^T = 0\}$$

The matrix  $H$  is a generator matrix of the dual code  $\mathcal{C}^\perp$  of  $\mathcal{C}$  for the usual scalar product.

## B. Shortened codes and punctured codes

### Definition 1. (Shortened code)[5]

Let  $\mathcal{C}$  be an  $[n, k, d]_{q^m}$ -linear code, choose a subset  $I \subset \{1, 2, \dots, n\}$  of coordinates such that  $|I| = i$  with  $1 \leq i \leq n$  and take the subcode of  $\mathcal{C}$  consisting of the codewords having 0 on those positions. Deleting the chosen coordinates in every codeword of this subcode yields a  $\mathbb{F}_{q^m}$ -linear code denoted  $Short_I(\mathcal{C})$ .  $Short_I(\mathcal{C})$  is called a shortened code of  $\mathcal{C}$  on  $I$ .

### Definition 2. (Punctured code)[5]

Let  $\mathcal{C}$  be an  $[n, k, d]_{q^m}$ -linear code, choose a subset  $I \subset \{1, 2, \dots, n\}$  of coordinates such that  $|I| = i$  with  $1 \leq i < d$ . Deleting the chosen coordinates in every codeword yields a  $\mathbb{F}_{q^m}$ -linear code denoted  $Punct_I(\mathcal{C})$ .  $Punct_I(\mathcal{C})$  is called a punctured code of  $\mathcal{C}$  on  $I$ .

If  $I = \{j\}$ , we denote  $Punct_I(\mathcal{C})$  by  $Punct_j(\mathcal{C})$  and  $Short_I(\mathcal{C})$  by  $Short_j(\mathcal{C})$  where  $j$ ,  $1 \leq j \leq n$ , is the deleted position.

For all vector  $x \in \mathbb{F}_{q^m}^n$ , we denoted by  $Punct_I(x) = (x_i)_{i \notin I}$ , if  $x$  is such that  $x_I = (x_i)_{i \in I} = 0$ , we denoted by  $Short_I(x) = (x_i)_{i \notin I}$ .

**Remark 1.** There are some links between shortening and puncturing operations. Indeed, let  $\mathcal{C}$  be an  $[n, k, d]_{q^m}$ -linear code. Let  $I$  be a subset of  $N = \{1, 2, \dots, n\}$ . The shortened code of  $\mathcal{C}$  on  $I$  is the punctured code on  $I$  of the subcode  $\mathcal{C}_I = \{c = (c_1, c_2, \dots, c_n) \in \mathcal{C} \mid c_i = 0 \forall i \in I\}$ . We remark also that a shortened code of a linear code  $\mathcal{C}$  can be considered like a subcode of  $\mathcal{C}$  if we replace the deleted coordinates by 0. Therefore, all the best decoding algorithms of  $\mathcal{C}$ , can be used to decode a shortened code of  $\mathcal{C}$ .

**Theorem 1.** Let  $\mathcal{C}$  be an  $[n, k, d]_{q^m}$ -linear code. Let  $I$  be a subset of  $N = \{1, 2, \dots, n\}$ . Then the following identity is verified

$$Punct_I(\mathcal{C})^\perp = Short_I(\mathcal{C}^\perp)$$

*Proof:*

Let  $x \in \mathcal{C}$ ,  $y \in \mathcal{C}^\perp$ . Then if  $(y_i)_{i \in I} = 0$ , we have  $Short_I(y) = (y_k)_{k \notin I} \in Short_I(\mathcal{C}^\perp)$  and  $Punct_I(x) = (x_k)_{k \notin I} \in Punct_I(\mathcal{C})$ . According to the definition of a code and its dual we have

$$\begin{aligned} x.y^T = 0 &\iff \sum_{k \in N} x_k y_k = \sum_{k \in N \setminus I} x_k y_k + \sum_{k \in I} x_k y_k = 0 \\ &\iff \sum_{k \in N \setminus I} x_k y_k = Punct_I(x).Short_I(y)^T = 0. \end{aligned}$$

■

**Lemma 1.** Let  $\mathcal{C}$  be an  $[n, k, d]_{q^m}$ -linear code. Let  $i \in \{1, \dots, n\}$ . The equality  $Short_i(\mathcal{C}) = Punct_i(\mathcal{C})$  is verified if and only if one of the following conditions is satisfied:

- 1) For all codewords  $c = (c_1, c_2, \dots, c_n) \in \mathcal{C}$ ,  $c_i = 0$ ,
- 2) The word  $e_i = (0, \dots, 0, 1, 0, \dots, 0)$  having only one non-zero coefficient which is equal to 1 in position  $i$  is in  $\mathcal{C}$ .

*Proof:*

- 1) Suppose first that the identity  $Short_i(\mathcal{C}) = Punct_i(\mathcal{C})$  is verified.  
Suppose that Conditions 1) is not satisfied, then there exists a codeword  $c \in \mathcal{C}$  such that  $c_i = 1$ . Under our hypothesis,  $Punct_i(c)$  is an element of  $Short_i(\mathcal{C})$ , i.e. there exists  $c' \in \mathcal{C}$  such that  $Punct_i(c) = Short_i(c')$ . Clearly,  $e_i = c' - c$  is an element of  $\mathcal{C}$  and Condition 2) is satisfied.
- 2) Reciprocally
  - a) Suppose that Condition 1) is satisfied. Since all the codewords  $c \in \mathcal{C}$  verify  $c_i = 0$ , then  $Punct_i(\mathcal{C}) = Short_i(\mathcal{C})$ .

- b) Suppose that Condition 2) is satisfied. Let  $\mathcal{C}_i = \{c \in \mathcal{C} \text{ s.t. } c_i = 0\}$  be the subcode of  $\mathcal{C}$  constituted of codewords  $c$  such that  $c_i = 0$ . Clearly,  $\mathcal{C}$  is generated by  $\{e_i\} \cup \mathcal{C}_i$ . Let  $c$  be a codeword of  $\mathcal{C}$ . If  $c \in \mathcal{C}_i$  then  $\text{Punct}_i(c) = \text{Short}_i(c) \in \text{Short}_i(\mathcal{C})$ . If  $c = e_i + c'$ ,  $c' \in \mathcal{C}_i$ , then  $\text{Punct}_i(c) = \text{Short}_i(c') \in \text{Short}_i(\mathcal{C})$ .

In both cases,  $\text{Short}_i(\mathcal{C}) = \text{Punct}_i(\mathcal{C})$ . ■

**Remark 2.** If a code  $\mathcal{C}$  satisfies the first condition of Lemma 1, then its dual  $\mathcal{C}^\perp$  will satisfy the second one.

We deduce the following proposition.

**Proposition 1.** Let  $\mathcal{C}$  be an  $[n, k, d]$   $\mathbb{F}_{q^m}$ -linear code and  $i$ ,  $1 \leq i \leq n$ , be an integer. If the parameters of  $\text{Punct}_i(\mathcal{C})$  and  $\text{Short}_i(\mathcal{C})$  are respectively  $[n-1, k_p, d_p]$  and  $[n-1, k_s, d_s]$ , then:

- 1)  $d_s \geq d$ ,  $d_s \geq d_p \geq d-1$ .
- 2) If  $\text{Punct}_i(\mathcal{C}) \neq \text{Short}_i(\mathcal{C})$  then  $k_p = k$  and  $k_s = k-1$ .
- 3) If  $\text{Punct}_i(\mathcal{C}) = \text{Short}_i(\mathcal{C})$  then
  - If Condition 1 of Lemma 1 is verified, then the parameters of  $\text{Punct}_i(\mathcal{C})$  and  $\text{Short}_i(\mathcal{C})$  are  $[n-1, k, d]$ .
  - If Condition 2 of Lemma 1 is verified (i.e.  $e_i \in \mathcal{C}$ ), then  $k_p = k_s = k-1$ .

*Proof:*

Note that, since  $\text{Short}_i(\mathcal{C}) \subset \text{Punct}_i(\mathcal{C})$ , the following relations hold:  $k_s \leq k_p \leq k$  and  $d_s \geq d_p$ . Moreover, using the notations of the proof of Lemma 1,  $\text{Short}_i(\mathcal{C})$  is isomorphic to  $\mathcal{C}_i \subset \mathcal{C}$ , and then  $d_s \geq d$ .

One can easily check that  $d_p \geq d-1$ .

Suppose firstly that for all codewords  $c = (c_1, c_2, \dots, c_n) \in \mathcal{C}$ ,  $c_i = 0$  (Condition 1 of Lemma 1). Then we have,  $\text{Punct}_i(\mathcal{C}) = \text{Short}_i(\mathcal{C})$  is an  $[n-1, k, d]$  code.

Suppose now that there exists a codeword  $c \in \mathcal{C}$  such that  $c_i = 1$ . One can check that the code  $\mathcal{C}$  is equal to  $\langle \{c\} \rangle \oplus \mathcal{C}_i$ . We deduce that  $k_s = k-1$ .

If  $e_i \notin \mathcal{C}$ , then  $\text{Short}_i(\mathcal{C}) \subsetneq \text{Punct}_i(\mathcal{C})$ , and then  $k_s = k-1 < k_p \leq k$ .

If  $e_i \in \mathcal{C}$ , then  $k_p = k_s = k-1$ . ■

One can notice that if  $e_i \in \mathcal{C}$ , then  $d = 1$ , and we have no information about values of  $d_s$  and  $d_p$  (but we have in this case  $d_s = d_p$ ).

From Proposition 1, we deduce the following corollary:

**Corollary 1.** Let  $\mathcal{C}$  be an  $[n, k, d]$   $\mathbb{F}_{q^m}$ -linear code and  $I$  be a set of  $r$  distinct positions. Let  $[n-1, k_s, d_s]$  and  $[n-1, k_p, d_p]$  be respectively the parameters of  $\text{Short}_I(\mathcal{C})$  and  $\text{Punct}_I(\mathcal{C})$ . Then we have  $d_s \geq d$ ,  $k_s \geq k-r$ ,  $d_p \geq d-r$  and  $k_p \geq k-r$ .

### C. Subfield Subcodes and Trace code

For more details and proofs, the reader can refer to [16], Ch.7 §7.

**Definition 3.** The subfield subcode  $\mathfrak{C}$  over  $\mathbb{F}_q$  of a  $\mathbb{F}_{q^m}$ -linear code  $\mathcal{C}$  is the set of codewords of  $\mathcal{C}$  which have components in  $\mathbb{F}_q$ :  $\mathfrak{C} = \mathcal{C} \cap \mathbb{F}_q^n$ .

A first property of  $\mathfrak{C}$  is the fact that it is a  $\mathbb{F}_q$ -linear code. The simplest way to construct such a subfield subcode is to construct a parity check matrix as follows.

Let  $\mathcal{C}$  be an  $[n, k, d]$   $\mathbb{F}_{q^m}$ -linear code defined by the parity check matrix

$$\mathbf{H} = \begin{pmatrix} h_{1,1} & \dots & h_{1,n} \\ \vdots & & \vdots \\ h_{r,1} & \dots & h_{r,n} \end{pmatrix} \in \mathbb{F}_{q^m}^{r \times n}.$$

Let  $\mathcal{B} = \{b_1, b_2, \dots, b_m\}$  be a basis of  $\mathbb{F}_{q^m}$  as a  $\mathbb{F}_q$ -vector space. We can construct the map  $\phi_{\mathcal{B}} : \mathbb{F}_{q^m} \rightarrow \mathbb{F}_q^m$  defined by, if  $x = \sum_{i=1}^m x_i b_i$ ,  $x_i \in \mathbb{F}_q$ , then  $\phi(x) = (x_1, x_2, \dots, x_m)$ .

**Proposition 2.** The matrix  $\tilde{\mathbf{H}} = \begin{pmatrix} \phi(h_{1,1})^T & \dots & \phi(h_{1,n})^T \\ \vdots & & \vdots \\ \phi(h_{r,1})^T & \dots & \phi(h_{r,n})^T \end{pmatrix}$  is a parity check matrix of the subfield subcode  $\mathfrak{C}$  of  $\mathcal{C}$ .

Note that  $\tilde{\mathbf{H}}$  is not necessary of full rank. Then a subfield subcode  $\mathfrak{C}$  of an  $[n, k, d]$   $\mathbb{F}_{q^m}$ -linear code  $\mathcal{C}$  is an  $[n, k^* \geq n - rm, d^* \geq d]$   $\mathbb{F}_q$ -linear code.

One can notice that  $\tilde{\mathbf{H}}$  is independent of the choice of the basis  $\mathcal{B}$ , which allows to omit the index  $\mathcal{B}$  in the definition of  $\phi_{\mathcal{B}}$ .

**Example 1.** The subfield subcode of the Reed-Solomon code  $RS_d$  of minimal distance  $d$  is the BCH code  $BCH_d$  of constructed minimal distance  $\delta = d$  over the prime subfield  $\mathbb{F}_p$ . Note that the true minimum distance of  $BCH_d$  could be greater than  $d$ .

Another construction of a  $\mathbb{F}_q$ -linear code from a  $\mathbb{F}_{q^m}$ -linear code is the trace construction.

If  $x$  is an element of  $\mathbb{F}_{q^m}$ , the trace of  $x$  over  $\mathbb{F}_q$  is defined by  $T_m(x) = x + x^q + x^{q^2} + \dots + x^{q^{m-1}}$ . The trace function is a  $\mathbb{F}_q$ -linear map. This mapping is naturally extended to  $\mathbb{F}_{q^m}^n$ : if  $c = (c_1, \dots, c_n) \in \mathbb{F}_{q^m}^n$ , then  $T_m(c) = (T_m(c_1), \dots, T_m(c_n)) \in \mathbb{F}_q^n$ .

**Definition 4.** [16] The trace code of an  $\mathbb{F}_{q^m}$ -linear code  $\mathcal{C}$  is the  $\mathbb{F}_q$ -linear code  $T_m(\mathcal{C})$ .

The link between trace code and subfield subcode is described in the following theorem:

**Theorem 2.** [16][Th.11, ch.7 §7, Delsarte] The dual of the subfield code  $\mathcal{C}$  of a code  $\mathcal{C}$  is the trace code of its dual:  
 $\mathcal{C}^\perp = T_m(\mathcal{C}^\perp)$ .

This fact is a direct consequence of Proposition 2 and a classical result of algebra: all  $\mathbb{F}_q$ -linear mapping of  $\mathbb{F}_{q^m}$  into  $\mathbb{F}_q$  can be expressed as  $T_m(\alpha x)$  for some  $\alpha \in \mathbb{F}_{q^m}$ .

#### D. $\mathbb{F}_{q^m}$ -linear isometries and Alternant codes

It is well-known [10] that the linear isometries for the Hamming distance on  $\mathbb{F}_{q^m}^n$  form a group

generated by the permutations of the support and the scalar multiplications by invertible elements of  $\mathbb{F}_{q^m}$  on each coordinate. From a matrix point of view, it is the monomial group  $\mathcal{M}_n$  of  $n \times n$  matrices over  $\mathbb{F}_{q^m}$  with one and only one non-zero element on each row and each column.

In order to obtain a code equivalent to  $\mathcal{C}$ , such a monomial matrix acts by right multiplication on any generator matrix of a code  $\mathcal{C}$ . The new code is  $\mathbb{F}_{q^m}$ -linear and has the same parameters of the original one.

Moreover, if  $\mathcal{C}$  has a decoding algorithm, then this algorithm can be used to decode the new code.

The most famous example is that of Generalized Reed-Solomon (GRS) codes that are obtained by applying a monomial matrix to a Reed-Solomon code.

An Alternant code is simply a subfield subcode of a GRS code. It naturally inherits the decoding algorithm of the underlying Reed-Solomon code.

#### E. $q$ -ary images of a code of length $n$ over $\mathbb{F}_{q^m}$

As we did in Section II-C, we fix a basis  $\mathcal{B} = (b_1, \dots, b_m)$  of  $\mathbb{F}_{q^m}$  over  $\mathbb{F}_q$  and denote by  $\phi_{\mathcal{B}}$  the corresponding  $\mathbb{F}_q$ -linear isomorphism  $\mathbb{F}_{q^m} \mapsto \mathbb{F}_q^m$ .

The mapping  $\phi_{\mathcal{B}}$  can be extended to the whole space  $\mathbb{F}_{q^m}^n$ : if  $c = (c_1, \dots, c_n) \in \mathbb{F}_{q^m}^n$ , then  $\Phi_{\mathcal{B}}(c) = (\phi_{\mathcal{B}}(c_1), \dots, \phi_{\mathcal{B}}(c_n))$ .

**Definition 5.** The  $q$ -ary image of a code  $\mathcal{C}$  relative to the base  $\mathcal{B}$  is the image  $Im_q(\mathcal{C}) = \Phi_{\mathcal{B}}(\mathcal{C})$  of  $\mathcal{C}$  by  $\Phi_{\mathcal{B}}$ .

The code  $Im_q(\mathcal{C})$  is clearly a  $\mathbb{F}_q$ -linear code of length  $nm$ . Note that, contrary to Section II-C, this code is dependent on the choice of the basis  $\mathcal{B}$ .

In order to build a generator matrix  $G$  of  $Im_q(\mathcal{C})$  over  $\mathbb{F}_q$  from those  $\mathcal{G}$  of  $\mathcal{C}$  over  $\mathbb{F}_{q^m}$ , since  $Im_q(\mathcal{C})$  is not  $\mathbb{F}_{q^m}$ -linear, it is necessary to take all the multiples of the rows of  $\mathcal{G}$ . In fact, it is sufficient to take  $m$  multiples  $\mathbb{F}_q$ -linearly independent.

A simple way is the following: for any element  $\beta \in \mathbb{F}_{q^m}$ , the map  $\psi_{\beta}: x \mapsto \beta x$  is a  $\mathbb{F}_q$ -linear endomorphism of  $\mathbb{F}_{q^m}$ . Its image by  $\phi_{\mathcal{B}}$  is an endomorphism of  $\mathbb{F}_q^m$ . We denote by  $M_{\beta}$  the matrix of the corresponding endomorphism: with obvious notations, if  $\phi_{\mathcal{B}}(x) = (x_1, \dots, x_m)$  then  $\phi_{\mathcal{B}}(\beta x) = (x_1, \dots, x_m)M_{\beta}$ .

**Proposition 3.** If  $\mathcal{G} = (\beta_{i,j})$  is a  $k \times n$  generator matrix of  $\mathcal{C}$ , then the  $mk \times nm$  matrix  $G$  obtained by replacing each entry  $\beta_{i,j}$  by the corresponding  $m \times m$  matrix  $M_{\beta_{i,j}}$ . Moreover, the matrix  $G$  is of full rank  $km$ .

*Proof:* The fact that  $G$  generates the full code  $Im_{\mathcal{B}}(\mathcal{C})$  comes directly from the fact that  $\Phi_{\mathcal{B}}$  is an isomorphism. In addition the two codes have the same number of elements, which implies that  $G$  is of rank  $km$ . ■

As a direct consequence, we obtain the following corollary:

**Corollary 2.** If  $\mathcal{C}$  is an  $[n, k, d]$   $\mathbb{F}_{q^m}$ -linear code, then  $Im_q(\mathcal{C})$  is an  $[nm, km, d_q \geq d]$   $\mathbb{F}_q$ -linear code.

### III. LINK BETWEEN SUBFIELD SUBCODES AND SHORTENED CODES

#### A. Shortening the $q$ -ary image of a code

Let  $u = (i_1, i_2, \dots, i_n) \in \{1, 2, \dots, m\}^n$  be a  $n$ -tuple of positions  $i_j$ ,  $1 \leq i_j \leq m$ . We define two sets of indexes  $I_u = \{i_1, i_2 + m, i_3 + 2m, \dots, i_n + (n-1)m\}$  and  $J_u = \overline{I_u} = \{1, \dots, n\} \setminus I_u$ .

Let  $\mathcal{C}$  be a linear code of length  $n$  over  $\mathbb{F}_{q^m}$ . We fix a basis  $\mathcal{B}$  of  $\mathbb{F}_{q^m}$  over  $\mathbb{F}_q$  and we look at the  $q$ -ary image of  $\mathcal{C}$  relatively to  $\mathcal{B}$ .

We denote by  $S_u$  (respectively  $P_u$ ) the operation of shortening (respectively puncturing) the  $q$ -ary image of  $\mathcal{C}$  on positions  $J_u$ :  $S_u(\mathcal{C}) = \text{Short}_{J_u}(\text{Im}_q(\mathcal{C}))$  and  $P_u(\mathcal{C}) = \text{Punct}_{J_u}(\text{Im}_q(\mathcal{C}))$ .

**Proposition 4.** *If  $\mathcal{C}$  is an  $[n, k, d]$   $\mathbb{F}_{q^m}$ -linear code, then  $S_u(\mathcal{C})$  is an  $[n, k', d']$   $\mathbb{F}_q$ -linear code with  $k' \geq n - m(n - k)$  and  $d' \geq d$ . Moreover, if the code  $\mathcal{C}$  has a decoding algorithm of error correction capability  $t$ , then this algorithm can be applied to  $S_u(\mathcal{C})$  with the same error correction capability.*

*Proof:* The inequalities  $k' \geq n - m(n - k)$  and  $d' \geq d$  are direct consequences of Corollary 1 and Corollary 2. In order to decode a noisy codeword  $y$  of  $S_u(\mathcal{C})$ , we extend  $y$  to a word of length  $nm$  by adding the value 0 on the shortened position, then we use the inverse of the map  $\Phi_B$  in order to obtain a noisy codeword  $\mathbf{y}$  of  $\mathcal{C}$ . By construction, the weight of the errors on  $y$  and  $\mathbf{y}$  are the same. So, if the error is less than or equal to  $t$ , it is possible to correct  $\mathbf{y}$  and to recover the correct codeword  $c \in S_u(\mathcal{C})$ . ■

**Example 2.** Set  $n = 7$ ,  $m = 3$  and let  $\alpha$  be a root of the polynomial  $x^3 + x + 1$ . The following matrix is a generator matrix of the Reed Solomon code  $RS_2$  of parameters  $[7, 6, 3]_8$  associated to the support  $a = (1, \alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6)$ :

$$\mathcal{G} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4 & \alpha^5 & \alpha^6 \\ 1 & \alpha^2 & \alpha^4 & \alpha^6 & \alpha & \alpha^3 & \alpha^5 \\ 1 & \alpha^3 & \alpha^6 & \alpha^2 & \alpha^5 & \alpha & \alpha^4 \\ 1 & \alpha^4 & \alpha & \alpha^5 & \alpha^2 & \alpha^6 & \alpha^3 \\ 1 & \alpha^5 & \alpha^3 & \alpha & \alpha^6 & \alpha^4 & \alpha^2 \end{pmatrix}$$

Its generator matrix in form systematic is given by:

$$\mathcal{G}_{sys} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \alpha \\ 0 & 1 & 0 & 0 & 0 & 0 & \alpha^2 \\ 0 & 0 & 1 & 0 & 0 & 0 & \alpha^3 \\ 0 & 0 & 0 & 1 & 0 & 0 & \alpha^4 \\ 0 & 0 & 0 & 0 & 1 & 0 & \alpha^5 \\ 0 & 0 & 0 & 0 & 0 & 1 & \alpha^6 \end{pmatrix}$$

The  $q$ -ary image (binary image) of the generator matrix  $\mathcal{G}$  in the base  $\{1 = (100), \alpha = (010), \alpha^2 = (001)\}$  is given by

$$\text{Im}_2(\mathcal{G}_{sys}) = \begin{pmatrix} M_1 & M_0 & M_0 & M_0 & M_0 & M_0 & M_\alpha \\ M_0 & M_1 & M_0 & M_0 & M_0 & M_0 & M_{\alpha^2} \\ M_0 & M_0 & M_1 & M_0 & M_0 & M_0 & M_{\alpha^3} \\ M_0 & M_0 & M_0 & M_1 & M_0 & M_0 & M_{\alpha^4} \\ M_0 & M_0 & M_0 & M_0 & M_1 & M_0 & M_{\alpha^5} \\ M_0 & M_0 & M_0 & M_0 & M_0 & M_1 & M_{\alpha^6} \end{pmatrix}$$

then

$$\text{Im}_2(\mathcal{G}) = \left( \begin{array}{ccc|ccc|ccc|ccc|ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{array} \right)$$

$$\text{with } \mathbf{M}_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{M}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{M}_{\alpha^i} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}^i \quad \text{for all } i \in \{1, 2, \dots, 6\}$$

The parity check matrix of the binary image  $Im_2(\mathcal{C})$  of the code  $\mathcal{C}$  is given by:

$$\mathcal{H}_2 = \left( \begin{array}{ccc|ccc|ccc|ccc|ccc|ccc} 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \end{array} \right)$$

Let  $u = (2, 3, 3, 2, 2, 3, 3) \in \{1, 2, 3\}^7$  be a tuple then  $I_u = \{2, 6, 9, 11, 14, 18, 21\}$  Now we compute  $S_u(\mathcal{H}_2)$  corresponding to the generator matrix of  $Im_2(\mathcal{C})$ :

$$S_u(\mathcal{H}_2) = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}$$

The generator matrix  $\mathcal{G}_{S_u}$  of the subfield subcode  $S_u(\mathcal{C})$  of the code  $\mathcal{C}$  over  $\mathbb{F}_2$  is given by

$$\mathcal{G}_{S_u} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Then  $S_u(\mathcal{C})$  is a  $[7, 4, 2]$  binary linear code.

When  $u = (1, 3, 1, 2, 3, 1, 3)$  we have  $I_u = \{1, 6, 7, 11, 15, 16, 21\}$  and the subfield subcode  $S_u(\mathcal{C})$  over  $\mathbb{F}_2$  of the code  $\mathcal{C}$  is an  $[7, 4, 3]$  binary linear code of generator matrix

$$\mathcal{G}_{S_u} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

Algorithm 1 give a simple method to construct a generator matrix  $G$  of  $S_u(\mathcal{C})$  from a generator matrix  $\mathcal{G}$  of  $\mathcal{C}$ .

---

**Algorithm 1** Generator matrix of  $G$  of  $S_u(\mathcal{C})$

---

- Construct a generator matrix of  $Im_q(\mathcal{C})$  using the method described in Section II-E.
  - Compute a generator matrix of the dual of this image.
  - Delete the columns indexed by  $J_u$  and perform a Gaussian elimination on this matrix.
  - Compute a generator matrix of the dual of this punctured code.
- 

### B. Subfield Subcode as shortened $q$ -ary image of a code

If we choose a basis  $\mathcal{B} = \{b_1, b_2, \dots, b_m\}$  of  $\mathbb{F}_{q^m}$  such that  $b_1 = 1$ , then we have  $\mathbb{F}_q = \phi^{-1}(\{(a_1, 0, 0, \dots, 0) \mid a_1 \in \mathbb{F}_q\})$ .

**Proposition 5.** Let  $u = (1, 1, \dots, 1)$  be a  $n$ -tuples of positions. If  $\mathcal{B}$  is a basis such that  $b_1 = 1$ , then the code  $S_u(\mathcal{C})$  is the subfield subcode of  $\mathcal{C}$  over  $\mathbb{F}_q$ .

*Proof:* This is a direct consequence of Remark 1: a codeword  $c = (c_1, \dots, c_n) \in \mathbb{F}_q^n$  is in  $S_u(\mathcal{C})$  if and only if  $c' = (c_1, 0, \dots, 0, c_2, 0, \dots, 0, \dots, c_n, 0, \dots, 0) \in \mathbb{F}_q^{nm}$  is in  $Im_q(\mathcal{C})$ , which is equivalent to the fact that  $\Phi_B^{-1}(c') \in \mathcal{C} \cap \mathbb{F}_q^n$ . ■

A natural question is: Does this construction allow to construct new codes? The answer will be given in the next section.

**Proposition 6.** Let  $u = (i, i, \dots, i)$  be a  $n$ -tuples of positions where  $1 \leq i \leq m$ . If  $\mathcal{B}$  is a multiplicative basis of the form  $\mathcal{B} = (1, \alpha, \alpha^2, \dots, \alpha^{m-1})$ , then the code  $S_u(\mathcal{C})$  is the subfield subcode over  $\mathbb{F}_q$  of  $\mathcal{C}$ .

We do not the give the proof of this result, but it will be easily derived from the discussion of the next section.

## IV. A FIRST ATTEMPT AT GENERALIZATION

In this section we will show that the codes introduced in Section III-A can also be constructed using the classical method used to construct alternant codes. However, our approach leads to a second generalization presented in Section V.

For an arbitrary finite field  $\mathbb{F}_q$  we denote by  $\mathbb{E} = \mathbb{F}_q^m$  and by  $GL_q(m)$  the linear group of isomorphisms acting on  $\mathbb{E}$ .



### A. $q$ -ary block codes of length $r$ over $\mathbb{E}$

For more details on block codes, the reader can refer to [1].

**Definition 6.** Let  $(A, +)$  be an additive group. An additive code of length  $r$  over  $A$  is an additive subgroup of  $(A^n, +)$ .

**Definition 7.** A block code of length  $n$  over  $\mathbb{E} = \mathbb{F}_q^m$  is an additive code over the additive group  $(\mathbb{E}^n, +)$  which is stable by scalar multiplication of any element  $\lambda$  of  $\mathbb{F}_q$ . The integer  $m$  is the size of the blocks.

Note that the condition on the scalar multiplication is not necessary if  $q = p$  is a prime number. Since  $\mathbb{E}^n$  is a  $\mathbb{F}_q$ -linear vector space of dimension  $nm$  isomorphic to  $\mathbb{F}_q^{nm}$ , a block code is also a  $\mathbb{F}_q$ -linear code of length  $nm$ . However, in this paper we are not interested in its properties as code of length  $nm$ , but in its block properties.

In particular, we look at its block-weight  $w_m$ , which denotes the number of non-zero blocks. For instance, the  $q$ -ary image  $Im_q(\mathcal{C})$  introduced previously is nothing else than a block code of size of blocks  $m$  and minimum block-distance equal to the minimum distance of  $\mathcal{C}$ .

Since a block code  $C$  is a  $\mathbb{F}_q$ -linear code, it is possible to define the notion of generator matrix, which is nothing else than the generator matrix of the corresponding linear code of length  $nm$  over  $\mathbb{F}_q$ . If its dimension is  $k$ , in order to compare a block code of block-size  $m$  with a  $\mathbb{F}_q^m$ -linear code, we introduce the notion of pseudo-dimension, which is  $k/m$ . Even if it is possible to construct the  $\mathbb{F}_q$ -dual of the linear code of length  $nm$ , the notion of duality for block code is not completely obvious, for example the  $q$ -ary image of the dual of a code  $\mathcal{C}$  over  $\mathbb{F}_q^m$  is not the dual of its  $q$ -ary image. More details on additive block codes, some generalizations of generator matrices and a notion of block-duality can be found in [1].

### B. Linear isometries of block codes

Generalizing the results of Section II-D, it is possible to define the isometry group of  $\mathbb{E}^n$ . We denote by  $w_m(x)$  the block weight of a word  $x \in E^n$ .

Let  $GL_q(m)$  be the group of  $\mathbb{F}_q$ -linear automorphism. Then  $GL_q(m)$  is isomorphic to the group of non-singular square matrices of length  $m \times m$  over  $\mathbb{F}_q$ .

For all  $f = (f_1, f_2, \dots, f_n) \in GL_q(m)^n$  and  $x = (x_1|x_2|\dots|x_n) \in \mathbb{E}^n$  where  $x_i \in \mathbb{E}$ , we define the action of  $GL_q(m)^n$  on  $E^n$  as follows:  $f(x) = (f_1(x_1)|f_2(x_2)|\dots|f_n(x_n))$ . This is equivalent to multiplying  $x$  by the block-diagonal matrix of size  $nm$  whose blocks are matrices of  $f_i$ .

**Theorem 3.** The  $\mathbb{F}_q$ -isometries of  $\mathbb{E}^n$  (i.e. linear isomorphisms preserving the Hamming block-weight) form a group generated by block permutations and  $GL_q(m)^n$ .

*Proof:* Let  $Mon_n(GL_q(m))$  be the group generated by the block permutations where each block is of length  $m$  and the block diagonal matrices of length  $nm$  which each matrix of length  $m$  on the diagonal is a non-singular matrix. It is clear that block permutations and the block diagonal matrices preserve the Hamming block weight of the element of  $\mathbb{E}^n$ .

Reciprocally, let  $g$  be an isometry of  $\mathbb{E}^n$ . We look at the images of elements of  $\mathbb{E}^n$  of block weight 1 by  $g$ . For  $1 \leq i \leq n$ , let  $V_i$  be the subspace of  $E^n$  of elements with all block component equal to 0 except the  $i$ -th: if  $x \in V_i$ , then  $x = (0, \dots, 0, x_i, 0, \dots, 0)$ ,  $x_i \in E$ . Pick an element  $x \in V_i$ . Since  $g$  is a block isometry,  $y = g(x) \in V_j$  for some  $j$ ,  $1 \leq j \leq n$ . Suppose that there exists another element  $x' \in V_i$  such that  $g(x') \in V_{j'}$ , with  $j \neq j'$ . Clearly  $w_m(x + x') = 1$  and  $w_m(g(x + x')) = 2$ . This implies that  $g(V_i) = V_j$ . So,  $g$  acts as a permutation on the set of  $V_i$ , which define the block-permutation part of our isometry. Applying the inverse of this permutation to  $g$ , we can now suppose that, for all  $i$ ,  $g(V_i) = V_i$ . If  $g_i$  denotes the restriction of  $g$  to  $V_i$ ,  $g_i$  must be  $\mathbb{F}_q$ -linear, moreover, since  $g_i$  preserves the block weight,  $Ker(g_i) = \{0\}$ , so  $g_i \in GL_q(m)$ , which completes the proof of this theorem. ■

The “monomial group”  $Mon_n(GL_q(m))$  introduced in the proof of Theorem 3 consists of the  $n \times n$  matrices having one and only one nonzero elements on each row and each column, moreover this non-zero element must be invertible and then is an element of  $GL_q(m)^n$ . So, this theorem is a generalization of Section II-D.

**Definition 8.** Let  $C$  and  $C'$  be two block codes of length  $n$  over  $E$ . The codes  $C$  and  $C'$  are equivalent if there exists an element  $f \in Mon_n(GL_q(m))$  such that  $C' = f(C)$ .

Clearly, if  $C' = f(C)$ , then the minimum block-distances of  $C$  and  $C'$  are equal. Moreover, if there exists a block-distance decoding algorithm for  $C$ , it can be used to decode  $C'$ .

There is no natural notion of duality for the block structure of a  $\mathbb{F}_q$ -linear code over  $E^n$ . However, we can look at the dual of a block code  $C$  considered as a code of length  $nm$  over  $\mathbb{F}_q$ .

If  $f_i \in GL_q(m)$  is a linear isomorphism, we denote by  $f_i^T$  its adjoint isomorphism. From a matrix point of view, this means that  $M_{f_i^T} = M_{f_i}^T$ .

**Proposition 7.** Let  $C$  be an additive code of length  $n$  over  $E$ ,  $f = (f_1, \dots, f_n) \in GL_q(m)^n$  be a diagonal isometry (without permutation) and  $C' = f(C)$ . Let  $f^* = ((f_1^{-1})^T, \dots, (f_n^{-1})^T) \in GL_q(m)^n$ . Then the relation between the dual of  $C$  and the dual of  $C'$  is  $C'^\perp = f^*(C^\perp)$ .

*Proof:* If  $x = (x_1, \dots, x_n) \in \mathbb{E}^n$  and  $y = (y_1, \dots, y_n) \in \mathbb{E}^n$ , then we have  $\langle x, y \rangle = \sum_{i=1}^n \langle x_i, y_i \rangle = \sum_{i=1}^n x_i y_i^T$ .

Applying this property to  $f(x)$  and  $f^*(y)$ , we obtain

$$\langle f(x), f^*(y) \rangle = \sum_{i=1}^n x_i M_{f_i} (y_i (M_{f_i}^{-1})^T)^T = \sum_{i=1}^n x_i M_{f_i} M_{f_i}^{-1} y_i^T = \sum_{i=1}^n x_i y_i^T = \langle x, y \rangle.$$

Consequently, we have  $\langle x, y \rangle = 0$  if and only if  $\langle f(x), f^*(y) \rangle = 0$ , which completes the proof. ■

In addition, it is easy to verify that the dual of a permuted block code is the permuted block code of its dual.

### C. Generalized Subfield Subcodes

Combining the results of Section III-A and Section IV-B, we are able to define the notion of generalized subfield subcode of a linear code  $\mathcal{C}$  over  $\mathbb{F}_{q^m}$ .

As before,  $\mathcal{B}$  denotes a fixed basis of  $\mathbb{F}_{q^m}$  over  $\mathbb{F}_q$ ,  $u$  denotes a set of indexes used to construct a shortened code,  $f = (f_1, \dots, f_n) \in GL_q(m)^n$  an  $n$ -tuple of linear isomorphisms,  $\pi$  a permutation of the  $n$  blocks and  $mon = \pi \circ f \in Mon_n(GL_q(m))$  the corresponding isometry.

**Definition 9.** Let  $\mathcal{C}$  be a  $\mathbb{F}_{q^m}$ -linear code of length  $n$ . The Generalized Subfield Subcode with relative to  $\mathcal{B}$ ,  $u$  and  $mon$  is the  $\mathbb{F}_q$ -linear code  $GSS(\mathcal{C}) = S_u(mon(Im_q(\mathcal{C})))$ .

We can immediately make some remarks.

#### Remark 3.

- If  $\mathcal{C}$  is an  $[n, k, d]_{q^m}$ -linear code, then  $GSS(\mathcal{C})$  is an  $[n, k' \geq n - m(n - k), d' \geq d]_q$ -linear code. Moreover, if  $\mathcal{C}$  has a decoding algorithm up to  $t$  errors, this algorithm can be applied to  $GSS(\mathcal{C})$ .
- In order to construct all the Generalized Subfield Subcodes of a code  $\mathcal{C}$ , it is not necessary to change the basis  $\mathcal{B}$ , indeed, a change of basis can be made by applying the corresponding matrix to the coordinates of  $f$ .
- In order to construct all the Generalized Subfield Subcodes of a code  $\mathcal{C}$ , it is sufficient to use the projections indexed by  $\bar{1} = (1, 1, \dots, 1)$ . Indeed, other values for the coordinates  $u_i$  correspond to permutations on each  $m$ -blocks, which is always a linear mapping in  $GL_q(m)$  and can be composed with the  $f_i$ 's.

Following Algorithm 1, Algorithm 2 allows to construct a generator matrix of a GSS code. The basis  $\mathcal{B}$  is fixed and  $u = (1, \dots, 1)$ .

---

#### Algorithm 2 Generator matrix of a GSS code

---

**Input:** A generator matrix  $\mathcal{G}$  of  $\mathcal{C}$  and  $mon \in Mon_n(GL_q(m))$

**Output:** A generator matrix  $G$  of  $GSS(\mathcal{C})$ , relative to  $mon$ .

- 1) Construct a generator matrix  $M$  of  $Im_q(\mathcal{C})$
- 2) Compute  $M' = M \text{Diag}_f$  where  $\text{Diag}_f$  is the  $nm \times nm$  block diagonal matrix, with each block corresponding to the  $f_i$ 's.
- 3) Compute a parity check matrix  $H'$  of  $M'$ .
- 4) Delete the columns of the matrix  $H'$  except the first ones of each block. This leads to a parity check matrix  $H$  of  $GSS(\mathcal{C})$ .
- 5) Perform a Gaussian elimination on  $H$ .
- 6) Permute  $H$  according to  $\pi$ .
- 7) Compute a parity check matrix  $G$  of  $H$

**Return:**  $G$

---

A first remark is the fact that the permutation  $\pi$  can be applied at any time from step 3 in the algorithm. However, it is simplest to perform the permutation at the end, since we no longer have to apply this permutation on blocks, but only on vectors of length  $n$ .

In addition, it is possible to use Proposition 7 in the algorithm by inverting the order of Step 2 and Step 3 and replacing  $f$  by  $f^*$ . This give the following variant for steps 2) to 4):

- 2) Compute a parity check matrix  $H$  of  $M$ .
- 3) Compute  $H' = H \text{Diag}_{f^*}$ .
- 4) Delete the columns of the matrix  $H'$  except the first ones of each block. This leads to a parity check matrix  $H$  of  $GSS(\mathcal{C})$ .

Let  $p_1: E \mapsto \mathbb{F}_q$  be the projection of an element on to its first component, the operations 2) and 3) of this variant can be combined into a single map  $p_1(f^*) = (p_1 \circ (f_1^{-1})^T, \dots, p_1 \circ (f_n^{-1})^T) \in (E^*)^n$  where  $E^*$  is the dual vector space of  $E$ , i.e.  $E^* = \mathcal{L}(E, \mathbb{F}_q)$ .



Remember that  $E^*$  is isomorphic to  $E$  as follows: for  $y \in E$ , we denote by  $\phi_y \in \mathbb{E}^*$  the map defined by  $\phi_y(x) = \langle x, y \rangle = xy^T$ . So, instead of choosing an element  $mon = \pi \circ f$  as input of Algorithm 2, we can choose a permutation  $\pi$  and an  $n$ -tuple  $y = (y_1, \dots, y_n) \in (E \setminus \{0\})^n$ . We denote by  $Diag_y$  the  $nm \times n$  block diagonal matrix with diagonal blocks  $y_i^T$ . Note that the diagonal blocks are not square matrices, but column vectors. So, the mapping  $p_1(f^*) = y$  is computed using  $Diag_y$ : for  $x = (x_1, \dots, x_n) \in E^n$ ,  $y(x) = (x_1 y_1^T, \dots, x_n y_n^T) = x Diag_y$ . This leads to Algorithm 3:

---

**Algorithm 3** Generator matrix  $G$  of  $GSS(\mathcal{C})$  relative to  $\pi$  and  $y$

---

**Input:** A generator matrix  $\mathcal{G}$  of  $\mathcal{C}$ , a permutation  $\pi$ , and a matrix  $Diag_y$  with  $y_i \neq 0$  for  $1 \leq i \leq n$ .

**Output:** A generator matrix  $G$  of  $GSS(\mathcal{C})$ , relative to  $\pi$  and  $y$ .

- 1) Construct a generator matrix  $M$  of  $Im_q(\mathcal{C})$
- 2) Compute a parity check matrix  $H$  of  $M$ .
- 3) Compute  $H' = M Diag_y$ .
- 4) Perform a Gaussian elimination on  $H'$ .
- 5) Permute  $H'$  according to  $\pi$ .
- 6) Compute a parity check matrix  $G$  of  $H'$

**Return:**  $G$

---

#### D. Link between generalized subfield subcodes and subfield subcodes of equivalent codes

In this section, we show that the generalized subfield subcodes of a given code are nothing else than subfield subcodes of equivalent codes. However, the approach presented in Section III gives a new point of view on this topic and will naturally be extended in the next section.

We need to have a more algebraic approach of the construction of generalized subfield subcodes. Suppose first without loss of generality, that the block permutation  $\pi$  is the identity. Indeed, this permutation can always be considered as having already been applied beforehand to the code  $\mathcal{C}$ .

We will look at a fixed coordinate of a word  $c = (c_1, \dots, c_n) \in \mathcal{C} \subset \mathbb{F}_{q^m}^n$ . We choose a coordinate  $u = c_i \in \mathbb{F}_{q^m}$ . Suppose that  $u = \sum_{i=1}^m u_i b_i$ ,  $u_i \in \mathbb{F}_q$  is the decomposition of  $u$  on the basis  $\mathcal{B}$ . Let  $M = M_{f_i}$  be a  $m \times m$  matrix corresponding to  $f_i \in GL_q(m)$ . This matrix  $M$  can be interpreted as a change of basis  $\mathcal{B}$  to  $\mathcal{B}' = (b'_1, \dots, b'_m)$ :  $(u_1, \dots, u_m)M$  is nothing else than the coordinates of  $u$  on this new basis  $\mathcal{B}'$ . Let  $V_i = V$  be the  $\mathbb{F}_q$ -subspace of  $\mathbb{F}_{q^m}$  generated by  $b'_1$ . The shortening operation in the  $i$ -th  $m$ -tuple in the construction of a generalized subfield subcode consists of keeping only the code words having their  $i$ -th coordinate in  $V_i$ , and then identify  $V_i$  to  $\mathbb{F}_q$  by means of its generator  $b'_1$ .

We have shown the following proposition:

**Proposition 8.** *The generalized subfield subcodes of a  $\mathbb{F}_{q^m}$ -linear code  $\mathcal{C}$  can be constructed as follows:*

- 1) Choose a set of  $n$   $\mathbb{F}_q$ -subspaces  $V_i$  of rank 1 of  $\mathbb{F}_{q^m}$ .
- 2) Set  $\mathcal{C}' = \mathcal{C} \cap \prod_{i=1}^n V_i$ .
- 3) By means of a generator  $a_i$  of each  $V_i$ , identifies  $V_i$  to  $\mathbb{F}_q$ . This leads to a  $q$ -ary image  $C = Im_q(\mathcal{C}')$ .
- 4) Choose a permutation  $\pi$  over  $\mathbb{F}_q^n$  and return  $\pi(C)$ .

One can remark that, since  $C$  is a  $\mathbb{F}_q$ -linear code, the construction does not depend on the choice of representatives  $a_i$ .

As a consequence of Proposition 8, we obtain the following theorem:

**Theorem 4.** *Let  $\mathcal{C}$  be an  $[n, k]_{q^m}$ -linear code. The generalized subfield subcodes of  $\mathcal{C}$  are exactly the codes obtained by taking the subfield subcodes of the  $\mathbb{F}_{q^m}$ -linear codes equivalent to  $\mathcal{C}$  under  $\mathbb{F}_{q^m}$ -linear isometries (as described in Section II-D).*

*Proof:*

Without loss of generality, we can suppose that, for both the generalized subfield subcode construction and the subfield subcode of equivalent codes construction, the permutation  $\pi$  is the identity.

Note that the subfield subcode of  $\mathcal{C}$  over  $\mathbb{F}_q$  corresponds to the choice  $V_1 = \dots = V_n = \mathcal{L}(1) = \mathbb{F}_q \subset \mathbb{F}_{q^m}$ .

Consider now any choice of subspaces  $V_i = \mathcal{L}(a_i)$ . Set  $D = Diag(a_1, \dots, a_n)$  be the  $n \times n$  diagonal matrix which corresponds to the multiplication of each component by the  $a_i$ 's. The subfield subcode of the image of  $\mathcal{C}$  by  $D$  is clearly the generalized

subfield subcode of  $\mathcal{C}$  corresponding to  $\prod_{i=1}^n V_i$ . ■

The following corollary is a direct consequence of Theorem 4.

**Corollary 3.** *The Generalized Subfield Subcodes of Reed-Solomon codes are exactly Alternant codes.*

In addition, we will make explicit the link between the subspaces  $V_i$  of Proposition 8 and the  $y_i$ 's of Algorithm 3.

**Proposition 9.** *The vector spaces of Proposition 8 are generated by the elements  $y_i$  of Algorithm 3.*

*Proof:* As previously, we denote by  $M = M_{f_i}$  the  $m \times m$  matrix corresponding to  $f_i \in GL_q(m)$ . This matrix  $M$  is interpreted as a change of basis  $\mathcal{B}$  to  $\mathcal{B}' = (b'_1, \dots, b'_m)$ . The matrix  $M^{-1}$  corresponds to the change of basis from  $\mathcal{B}'$  to  $\mathcal{B}$ . Its first row is given by the coordinates of  $b'_1$  in the basis  $\mathcal{B}$ . In the construction of Algorithm 3, the coordinates of  $y_i$  are given by the first column of  $(M^{-1})^T$ . Consequently, we have  $b'_1 = a_i = y_i$ , which completes our proof. ■

## V. SUBSPACE SUBCODES

Codes defined over a finite field of great size are used to correct burst errors or for concatenation of codes. The most famous example is that of Reed-Solomon codes over  $\mathbb{F}_{2^m}$ , with typical values  $4 \leq m \leq 8$ . However Reed-Solomon codes are MDS codes, which implies in particular that their length  $n$  is limited to  $2^m$ .

In practice, for transmission applications, a code over  $\mathbb{F}_{2^m}$  is generally implemented in binary, *i.e.* using its binary image  $Im_2(\mathcal{C})$ . The notion of additive block codes over  $E = \mathbb{F}_2^m$  is an interesting and efficient alternative for applications. In this section, we will present a generic construction of additive block codes of length greater than  $2^m$ . If the starting code is a Reed-Solomon code, these new codes possess a decoding algorithm and have a constructed minimal distance, which remains very competitive even if these codes cannot be MDS.

In this section, we introduce a new class of additive block codes with interesting parameters for both transmission and cryptography applications.

In order to facilitate a comparison between the parameters of linear codes over  $\mathbb{F}_{q^m}$  and block codes over  $\mathbb{F}_q^m$ , we use the notation  $[n; k; d]_{q^m}$  for their parameters, with  $n$  is the  $m$ -block length of the code,  $k = \text{Log}_{q^m}(\#\mathcal{C})$  is its pseudo-dimension and  $d$  is its  $m$ -block minimum distance. Note that, for an additive block code,  $k$  is not necessarily an integer.

In order to simplify the presentation of this section, we do not discuss the presence of a possible permutation  $\pi$  which is implicitly fixed to be the identity.

### A. Subspace subcodes

A natural, simple and efficient way to generalize the approach introduced in Section IV-D is to increase the size of the subspaces  $V_i$ .

**Definition 10.** *Let  $\mathcal{C}$  be an  $\mathbb{F}_{q^m}$ -linear code of length  $n$  and  $V$  be a  $\mathbb{F}_q$ -subspace of  $\mathbb{F}_{q^m}$  of dimension  $r \leq m$ . The subspace subcode over  $V$  of  $\mathcal{C}$  is the  $\mathbb{F}_q$ -linear code  $SS_V(\mathcal{C}) = \mathcal{C} \cap V^n$ .*

Most of the previous results can be generalized directly. Fixing a basis  $\mathcal{B} = (\beta_1, \dots, \beta_r)$  of  $V$ , the code  $SS_V(\mathcal{C})$  can be identified by an additive block code over  $\mathcal{E} = \mathbb{F}_q^r$ . If we complete the basis  $\mathcal{B}$  into a basis  $\mathcal{B}$  of  $E$ , this block code is obtained by shortening the  $q$ -ary image of  $\mathcal{C}$  over the  $m - r$  last components of each block.

We deduce directly the following proposition:

**Proposition 10.** *If the parameters of  $\mathcal{C}$  are  $[n, k, d]_{q^m}$ , then those of  $SS_V(\mathcal{C})$  are  $[n, k' \geq (km - n(m - r)), d' \geq d]_{q^r}$ .*

Note that, if we choose another basis  $\mathcal{B}$  of  $V$ , it leads to an equivalent (in the meaning of Section IV-B) block code.

In addition, if there is a decoding algorithm for  $\mathcal{C}$ , this algorithm can be applied to decode  $SS_V(\mathcal{C})$ .

### B. Generalized subspace subcode

**Definition 11.** *Let  $\mathcal{C}$  be a  $\mathbb{F}_{q^m}$ -linear code of parameters  $[n; k; d]_{q^m}$ . Let  $r$  be an integer less than  $m$ . Let  $V_1, \dots, V_n$  be a set of  $n$   $\mathbb{F}_q$ -subspaces of  $\mathbb{F}_{q^m}$  of dimension  $r$ . Set  $W = \prod_{i=1}^n V_i$ , constituted of  $n$ -tuples with the  $i$ -th coordinate in  $V_i$ . The generalized subspace subcode of  $\mathcal{C}$  relative to  $W$  is the  $\mathbb{F}_q$ -vector space  $GSS_W(\mathcal{C}) = \mathcal{C} \cap W$ .*

Proposition 10 can be also applied to generalized subspace subcodes.

**Example 3.** *Following Example 2, we start from the Reed Solomon code  $RS_3$  with parameters  $[7; 5; 3]_8$ . A generator matrix of the dual of its binary image is*

$$\mathcal{H} = \left( \begin{array}{ccc|ccc|ccc|ccc|ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{array} \right)$$

We choose  $W = V_1 V_2 V_1 V_3 V_1 V_2 V_1$ , where  $V_1$  is generated by 1 and  $\alpha$ ,  $V_2$  by 1 and  $\alpha^2$  and  $V_3$  by  $\alpha$  and  $\alpha^2$ . So, in order to obtain a parity check matrix of  $\mathcal{C} = GSS_W(RS_3)$ , we delete the columns indexed by 3, 5, 9, 10, 15, 17 and 21 of  $\mathcal{H}$ . A binary generator matrix of  $\mathcal{C}$  is then

$$\mathcal{G} = \left( \begin{array}{c|c|c|c|c|c|c|c|c|c|c|c} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{array} \right)$$

As a binary code, its parameters are  $[14; 8; 3]_2$ . If we look at this code as a block-code of block size equals to 2, it is a  $[7; 4; 3]_4$  code, which is optimal compared to a code over  $\mathbb{F}_4$  of length 7 and dimension 4.

In practice, if we want to construct such codes, it is easy to extend the previous algorithms to Algorithm 4.

---

**Algorithm 4** Generator matrix  $G$  of  $GSS_W(\mathcal{C})$

---

**Input:** A generator matrix  $\mathcal{G}$  of  $\mathcal{C}$ , a set  $W = \prod_{i=1}^n V_i$  of  $n$  vector spaces  $V_i$  of dimension  $r$ . Each  $V_i$  is defined by a basis  $(v_{i,1}, \dots, v_{i,r})$  of  $V_i$ .

**Output:** A generator matrix  $G$  of  $GSS_W(\mathcal{C})$ .

- 1) Construct a generator matrix  $M$  of  $Im_q(\mathcal{C})$
- 2) Construct  $n$  matrices  $M_i$  of size  $r \times m$ . The rows of  $M_i$  are the coordinates of the basis  $(v_{i,1}, \dots, v_{i,r})$  of  $V_i$  relative to  $\mathcal{B}$ .
- 3) Compute  $D = \text{Diag}(M_1^T, \dots, M_n^T)$ , the block-diagonal matrix of size  $mn \times rn$  having the matrices  $M_i^T$  as diagonal blocks.
- 4) Compute  $H' = HD$ .
- 5) Perform a Gaussian elimination on  $H'$ .
- 6) Compute a parity check matrix  $G$  of  $H'$

**Return:**  $G$

---

### C. Examples

In this construction, the minimum distance of the original code is preserved, while the dimension of the code decreases with the number of punctured positions. So, the value  $r = m - 1$  seems interesting to provide codes with nice parameters.

We give some examples.

- $q = 2, m = 4, r = 3$ . We start from the extended Reed-Solomon code  $\mathcal{C}$  over  $\mathbb{F}_{16}$  with parameters  $[16; 13; 4]_{16}$ . For  $r = 3$ , the parameters of any generalized subspace subcode of  $\mathcal{C}$  are  $[16; k' \geq 12; d' \geq 4]_8$ . Note that the parameters  $[16; 12; 4]_8$  are optimal for  $\mathbb{F}_8$ -linear code.  
In practice, all the codes we obtained had parameters exactly  $[16; 12; 4]_8$ .
- $q = 2, m = 5, r = 3$ . We start from the extended Reed-Solomon code  $\mathcal{C}$  over  $\mathbb{F}_{32}$  with parameters  $[32; 26; 7]_{16}$ . For  $r = 3$ , the parameters of any generalized subspace subcode of  $\mathcal{C}$  are  $[32; k' \geq 22; d' \geq 7]_8$ . The parameters  $[32; 22; 7]_8$  are optimal for  $\mathbb{F}_8$ -linear code.
- $q = 2, m = 9$ . We start from the extended Reed-Solomon code  $\mathcal{C}$  over  $\mathbb{F}_{29}$  with parameters  $[512; 350; 163]_{512}$ . For  $r = 83$ , the parameters of any generalized subspace subcode of  $\mathcal{C}$  are  $[512; k' \geq 329.75; d' \geq 163]_{256}$ .

### D. Cryptographic applications

The purpose of this section is not to present the complete design of a public key cryptosystem, but to show that the generalized subspace subcode construction is a promising research direction to hide the structure of a code.

The general principle of such a cryptosystem is as follows: the starting point is a class of codes for which there exists an efficient decoding algorithm up to a fixed number  $t$  of errors. The structure of such a code is masked by some operations which constitute the secret key. The public key is then a generator matrix of a code which looks like a random code  $\mathcal{C}$ . The message to be encrypted is encoded by a generator matrix of  $\mathcal{C}$  and a random error of weight  $t$  is added to this message.

Such a cryptosystem is sensitive to two types of attacks:

- Structural attacks that involve retrieving the structure of the masked code.

- Decoding by brute force. This consists of applying generic decoding algorithms to a random code. This problem is NP-hard, however the parameters of the code must be sufficiently large to resist at this kind of attacks.

The evaluation of the brute force decoding attack is not easy, and many papers are devoted to this topic [4], [8], [11], [19]. We chose to use a simple criterion: for a given code of parameters  $[n; k; d]$  with correction capacity  $t = \lfloor (d-1)/2 \rfloor$ , we compute the ratio between the number of information sets and the number of information sets without errors. Our measure of the workfactor is then  $wf = \binom{n}{k} / \binom{n-t}{k}$ . Our criterion yields a workfactor greater or equal to  $2^{128}$ .

Most of McEliece-like cryptosystems use subfield subcodes of Generalized Reed-Solomon codes (for example the binary Goppa codes in the original McEliece cryptosystem). We propose to use generalized subspace subcodes of Reed-Solomon codes (GSS codes of GRS codes for short).

There are some advantages and disadvantages to using generalized subspace subcodes instead of subfield subcodes.

- In the case  $r = 1$  and  $q = 2$ , our generalized subspace subcodes are nothing else than Alternant codes, and it is well-known that Goppa codes have better parameters.
- For  $r$  close to  $m$  (typically  $m-4 \leq rm$ ), the code parameters become more interesting. Moreover, if we want to construct subfield subcodes of a code over  $\mathbb{F}_{q^m}$ , the size of subfield is bounded above by  $2^{m'}$ , with  $m' \leq m/2$ . So our parameters are more flexible. Finally, there exist some attacks against subfield subcodes of GRS codes over large fields [6], [7]. A priori, this type of attack does not apply to generalized subfield subcodes.
- The main reason for this comes from the fact that the GSS codes are no longer defined over a field but over some vector space. In return, the description of these codes as linear codes over  $\mathbb{F}_{2^{m'}}$  cannot be used. We need to give a full  $\mathbb{F}_q$  basis of our codes, which increase the size of the secret key.

In practice, a binary Goppa code of parameters  $[4096; 3556; 91]$  leads to a resistance against brute force decoding greater than our criterion  $wf \geq 2^{128}$ . The corresponding size of the public key is then 938 Ko.

Our third example in Section V-C has for parameters  $[512; 329; 163]_{28}$  and leads to a workfactor greater than  $2^{128}$ . In addition, one can notice that we did not take in account the fact that this code is defined over a large alphabet, which will increase the complexity of this attack. Unfortunately, the size of the secret key is very large, approximatively 1514 Ko. This is essentially due to the fact that  $q = 2$ , that implies the code is  $\mathbb{F}_2$ -linear.

An intermediate solution consists in choosing a relatively large  $q$  and a small  $m$ . Set  $q = 2^4 = 16$ ,  $m = 3$  and  $r = 2$ . We obtain  $\mathbb{F}_{q^m} = \mathbb{F}_{2^{12}}$ , so it is possible to pick a Reed-Solomon code up to the length  $2^{12} = 4096$ . For example, we choose a Reed-Solomon code of parameters  $[700; 580; 121]_{2^{12}}$ . The  $V_i$  subspaces of our construction are subspaces over  $\mathbb{F}_{16}$  of dimension 2. We obtain a  $\mathbb{F}_{16}$ -linear GSS code of parameters  $[700; 520; 121]_{28}$ , which leads to a workfactor  $wf$  greater than  $2^{128}$ . The  $\mathbb{F}_q$ -generator matrix is of size  $1400 \times 1040$ . Each entry is over  $\mathbb{F}_{2^4}$  and needs 4 bits of memory. As usual, we use a systematic matrix to describe our code *i.e.* a matrix of size  $k \times n - k$ . The size of the public key is then  $1040 \times 360 \times 4 = 1497600$  bits, or 183 Ko, which is significantly smaller than a classical Goppa code with the same level of security against brute force decoding.

## VI. CONCLUSION

The purpose of this paper is not to present a complete study of structural attacks against subspace subcodes. Here is a list of questions that naturally come to mind and deserve further development.

- 1) The equivalent of GRS codes in the subspace subcode context correspond to generalized subspace subcodes of Reed-Solomon code with parameter  $r = m$ . It is well known that, from a generator matrix of a GRS code, it is easy to recover the underlying algebraic structure, *i.e.* the support of the corresponding Reed-Solomon code and the values of the scalar multiplications on each component [18].  
The problem in the GSS context is the following: we fix a Reed-Solomon code  $\mathcal{C}$  of parameters  $[n; k; d]$  over  $\mathbb{F}_{q^m}$ . We choose a basis  $\mathcal{B}$  and compute  $Im_q(\mathcal{C})$ . We pick randomly an element  $mon$  in  $Mon_n(GL_q(m))$  and compute a  $\mathbb{F}_q$ -generator matrix  $G$  of  $mon(Im_q(\mathcal{C}))$ . From the matrix  $G$ , is it possible in reasonable time to recover  $RS$  and  $mon$  or another equivalent set of parameters  $RS'$  and  $mon'$ ?
- 2) Given a Reed-Solomon code  $\mathcal{C}$  and a  $\mathbb{F}_q$ -generator matrix under systematic form of a generalized subspace subcode  $GSS_W(\mathcal{C})$ , is it possible to recover the secret basis of the subspaces  $V_i$  and the permutation  $\pi$ ? This question is connected to a list of problems in increasing order of difficulty:
  - Suppose  $\pi$  is known (which equivalent to  $\pi$  is the identity).
    - $r = m$ . This particular case will probably be solved using the conjugacy of matrices.
    - $1 \leq r < m$ .
  - $\pi$  is unknown.
    - $r = m$ .
    - $1 \leq r < m$ .
- 3) Given a  $\mathbb{F}_q$ -generator matrix under systematic form of a generalized subspace subcode  $GSS_W(\mathcal{C})$  with  $\mathcal{C}$  an unknown Reed-Solomon code, is it possible to recover  $\mathcal{C}$  and the algebraic parameters of  $GSS_W(\mathcal{C})$ ?

## REFERENCES

- [1] T. P. Berger and N. E. Amrani, "Codes over  $L(\text{GF}(2^m), \text{GF}(2^m))$ , MDS diffusion matrices and cryptographic applications," in *Codes, Cryptology, and Information Security - C2SI 2015, Proceedings*, ser. Lecture Notes in Computer Science, S. E. Hajji, A. Nitaj, C. Carlet, and E. M. Souidi, Eds., vol. 9084. Springer, 2015, pp. 197–214.
- [2] T. P. Berger, P. Cayrel, P. Gaborit, and A. Otmani, "Reducing key length of the McEliece cryptosystem," in *Progress in Cryptology - AFRICACRYPT Proceedings*, ser. Lecture Notes in Computer Science, B. Preneel, Ed., vol. 5580. Springer, 2009, pp. 77–97.
- [3] T. P. Berger and P. Loidreau, "How to mask the structure of codes for a cryptographic use," *Des. Codes Cryptography*, vol. 35, no. 1, pp. 63–79, 2005.
- [4] D. J. Bernstein, T. Lange, and C. Peters, "Smaller decoding exponents: Ball-collision decoding," in *Advances in Cryptology - CRYPTO 2011, Proceedings*, ser. Lecture Notes in Computer Science, P. Rogaway, Ed., vol. 6841. Springer, 2011, pp. 743–760.
- [5] G. Cohen, I. Honkala, S. Lytsin and A. Lobstein, "Covering codes" North Holland Mathematical Library, 1997.
- [6] J. Faugère, A. Otmani, L. Perret, F. de Portzamparc, and J. Tillich, "Structural cryptanalysis of McEliece schemes with compact keys," *Des. Codes Cryptography*, vol. 79, no. 1, pp. 87–112, 2016.
- [7] J. Faugère, A. Otmani, L. Perret, and J. Tillich, "Algebraic cryptanalysis of McEliece variants with compact keys," in *Advances in Cryptology - EUROCRYPT 2010, Proceedings*, ser. Lecture Notes in Computer Science, H. Gilbert, Ed., vol. 6110. Springer, 2010, pp. 279–298.
- [8] M. Finiasz and N. Sendrier, "Security bounds for the design of code-based cryptosystems," in *Advances in Cryptology - ASIACRYPT 2009, Proceedings*, ser. Lecture Notes in Computer Science, M. Matsui, Ed., vol. 5912. Springer, 2009, pp. 88–105.
- [9] P. Gaborit, "Shorter keys for code based cryptography" in *International Workshop on Coding and Cryptography - WCC 2005, Proceedings*, Bergen, Norway, Mar. 2005, pp. 81–91.
- [10] W.C. Huffman, "Groups and codes" in V.S. Pless and W.C. Huffman, editors, *Handbook of Coding Theory*, chapter 17. Elsevier, Amsterdam, The Netherlands, 1998.
- [11] P. J. Lee and E. F. Brickell, "An observation on the security of McEliece's public-key cryptosystem," in *Advances in Cryptology - EUROCRYPT '88, Proceedings*, ser. Lecture Notes in Computer Science, C. G. Günther, Ed., vol. 330. Springer, 1988, pp. 275–280.
- [12] R. Misoczki and P. S. L. M. Barreto, "Compact McEliece keys from goppa codes," in *Selected Areas in Cryptography - SAC 2009, Revised Selected Papers*, ser. Lecture Notes in Computer Science, M. J. J. Jr., V. Rijmen, and R. Safavi-Naini, Eds., vol. 5867. Springer, 2009, pp. 376–392.
- [13] R. Misoczki, J. Tillich, N. Sendrier, and P. S. L. M. Barreto, "MDPC-McEliece: New McEliece variants from moderate density parity-check codes," in *Proceedings of the 2013 IEEE International Symposium on Information Theory*. IEEE, 2013, pp. 2069–2073.
- [14] P. S. L. M. Barreto, R. Lindner, and R. Misoczki, "Monoidic codes in cryptography," in *Post-Quantum Cryptography - PQCrypto 2011, Proceedings*, ser. Lecture Notes in Computer Science, B. Yang, Ed., vol. 7071. Springer, 2011, pp. 179–199.
- [15] R. McEliece, "A public-key cryptosystem based on algebraic coding theory" DSN Prog. Rep., Jet Prop. Lab., California Inst. Technol., Pasadena, CA (January 1978) pp. 114–116.
- [16] F.J. MacWilliams and N.J.A. Sloane, "The Theory of Error Correcting Codes" North-Holland, Amsterdam, 1986.
- [17] E. Persichetti, "Compact McEliece keys based on quasi-dyadic Srivastava codes," *J. Mathematical Cryptology*, vol. 6, no. 2, pp. 149–169, 2012. [
- [18] V. M. Sidel'nikov and S. O. Shestakov, "On cryptosystems based on generalized Reed-Solomon codes," *Discrete Mathematics*, vol. 4, no. 3, pp. 57–63, 1992.
- [19] R. C. Torres and N. Sendrier, "Analysis of information set decoding for a sub-linear error weight," in *Post-Quantum Cryptography - PQCrypto 2016, Proceedings*, ser. Lecture Notes in Computer Science, T. Takagi, Ed., vol. 9606. Springer, 2016, pp. 144–161.
- [20] C. Wieschebrink, "Cryptanalysis of the Niederreiter public key scheme based on GRS subcodes," in *Post-Quantum Cryptography, PQCrypto 2010, Proceedings*, ser. Lecture Notes in Computer Science, N. Sendrier, Ed., vol. 6061. Springer, 2010, pp. 61–72.